DOCUMENT RESUME

ED 193 319 TM 800 636

AUTHOR Nimmer, Donald N.
TITLE The Use of Standardized Achievement Test Batteries in the Evaluation of Curriculum Changes in Junior High School Earth Science.
PUB DATE [80]
NOTE 13p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Tests; *Curriculum Evaluation; *Earth Science; Junior High Schools; Science Curriculum; *Secondary School Science; *Standardized Tests; *Test Interpretation; Test Items
IDENTIFIERS Comprehensive Tests of Basic Skills; Metropolitan Achievement Tests; Sequential Tests of Educational Progress; SRA Achievement Series; Stanford Achievement Tests

ABSTRACT

The sampling of test items found within the Science portion of selected standardized achievement test batteries is examined. The lack of sensitivity of such items to the curriculum changes often attempted in Earth Science is demonstrated. The following five standardized achievement test batteries commonly administered to pupils in grades 7-9 were selected for use in this study: Comprehensive Tests of Basic Skills, Metropolitan Achievement Tests, SRA Achievement Tests, Sequential Tests of Educational Progress (STEP), and Stanford Achievement Tests. Individual test items were classified into the general categories of Earth Science, Life Science, and Physical Science. Test items within general categories were further subdivided into content areas. With the exception of STEP, the three sciences were not represented within a given achievement test battery by equal numbers of test items. In addition, some of the achievement test batteries omitted or undersampled items in one or more of the content areas within the Earth Sciences. It is imperative to convince decision makers that standardized achievement test batteries should not be used to evaluate curriculum changes. (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED193319

The Use of Standardized Achievement Test Batteries in the Evaluation of Curriculum Changes in Junior High School Earth Science

Donald N. Nimmer

Director of Testing

Central Missouri State University

Warrensburg, Missouri

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. Nimmer

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 800636

Abstract

With the decrease of financial resources available for the implementation of changes to the Earth Science curricula in junior high schools, many local school districts have resorted to the use of the scores earned on the Science portion of the annually administered standardized achievement test batteries as a means of evaluating the relative effectiveness of those changes. The purpose of this study is to examine the sampling of test items found within the Science portion of selected standardized achievement test batteries and to demonstrate their lack of sensitivity to the curriculum changes often attempted in Earth Science.

The Use of Standardized Achievement Test Batteries in the Evaluation of Curriculum Changes in Junior High School Earth Science

Introduction

From the time the Soviets launched Sputnik in October, 1957, until the late 1960's, virtually unlimited funds were made available for the development, purchase, and implementation of new science curricula by local school districts. Financial support became available from numerous sources: public and private, business and government, and local, state, and federal agencies. So much support was available for new science curricula that the local classroom teacher had but to ask in order to receive. Little demonstration of the relative effectiveness of these programs was necessary since the money and public opinion in support of the sciences was in abundance.

The decade of the 1970's saw a reversal of this trend; fewer funds were available and public support for the sciences in general, and science curricula in particular, waned. Consequently, the competition for federal and state funds became keener and the application procedures became increasingly complex and sophisticated. Funding agencies began to require formal plans by which to evaluate the effectiveness of the curricula, innovations, and pedagogies being attempted. Accordingly, the funds available from within the local school district became more difficult to obtain as the academic priorities shifted from the sciences and fewer dollars were available for the purchase of new curricula and accompanying equipment.

The presence of a procedure by which to evaluate the outcomes of a specific program became so important that funding agencies typically

allowed and even encouraged the expenditure of a portion of the total budget for the employment of a professional evaluator and for the design, selection, purchase, and analysis of formal evaluation instruments. Therefore, while some and even many of the resultant evaluation components of projects funded with state and federal monies may have been open to criticism, the funds were available for the implementation of sound evaluation practices. In contrast, the same was usually not true for funds originating from within the budget of the local school district; most school districts are small and do not have access to the services of a professional evaluator.

Frequently the evaluation process is determined by the constraints imposed by the building principal, district superintendent, or other local officials having little or no expertise in the techniques used for the evaluation of curricula. These individuals are reluctant to commit the resources of the school district to the widespread infusion of a curricular change unless that change can be demonstrated to be effective.

The Problem

Each year local school districts administer standardized achievement test batteries to their pupils. One of the functions of the resultant scores is to provide a comparison between the performance of their pupils, and thereby the programs to which they have been exposed, and the performance of similar pupils from throughout the local school district, state, region, and nation. These scores represent a considerable expense in actual dollars as well as student and staff time.

School officials, being pragmatic by nature, often view the results obtained by the annual administration of standardized achievement test

batteries as a means of evaluating any revisions to the curriculum. After all, it is reasoned, these tests have been administered for many years so they provide a good longitudinal baseline from which to evaluate any changes. Furthermore, as long as the district has already allocated time and money for these tests, why not use the resultant scores? Pupils take too many tests already--why add another only to assess the effectiveness of a given curricular change? This type of logic, considering the level of curriculum evaluation expertise possessed by school officials, is difficult to fault--the currently administered achievement test batteries can provide them with a yardstick by which to measure achievement. Unfortunately, as it will be demonstrated for the area of Earth Science in the junior high school, standardized achievement test batteries are not sensitive enough to use for the evaluation of changes in the curriculum.

Data

Five standardized achievement test batteries commonly administered to pupils of junior high school age (grades 7-9) were selected for use in this study: the Comprehensive Tests of Basic Skills (Levels 3 and 4), the Metropolitan Achievement Tests (Advanced 1 Level), the SRA Achievement Tests (Levels F and G), the Sequential Tests of Educational Progress (Advanced 1 Level), and the Stanford Achievement Tests (Advanced Level). The publisher, date of publication, level designations, and the grade levels of these tests are provided in Table 1. The individual test items comprising the Science portion of these achievement test batteries were classified into the broad categories of Earth Science, Life Science, and Physical Science. The results of this breakdown are likewise reported

in Table 1.

- - - - - - - - - - -

Table 1

- - - - - - - - - - -

A perusal of Table 1 will reveal that, with the exception of the Sequential Tests of Educational Progress, the three sciences are not represented within a given achievement test battery by equal numbers of test items. The relative amount of emphasis for the Earth Sciences, for example, ranges from 33 1/3% (25 items) of the total Science portion of the Sequential Tests of Educational Progress (Advanced I Level) to only 7 1/2% (3 items) of the total Science portion of the Comprehensive Tests of Basic Skills (Level 4). An implication of this imbalance readily becomes apparent when the curriculum emphasis of the science program within a given junior high school differs from the balance of test items found in the achievement test battery: the greater the disparity, the more difficult it becomes to accurately ascertain how much has been learned within the science curriculum during a given period of time, grade, and/or course.

The test items within the general categories of Earth Science, Life Science, and Physical Science may be further subdivided into content areas. To illustrate, Earth Science is typically subdivided into Astronomy, Geology, Meteorology, and Oceanography. The data reported in Table 2 provides a breakdown of the test items in Earth Science

- - - - - - - - - - -

Table 2

- - - - - - - - - - -

From an inspection of the data appearing in Table 2 it would seem that some of the achievement test batteries have omitted or undersampled items in one or more of the content areas within the Earth Sciences. The Comprehensive Tests of Basic Skills (Levels 3 & 4), the SRA Achievement Tests (Level F), and the Stanford Achievement Tests (Advanced Level) have no questions on Oceanography while the Metropolitan Achievement Tests (Advanced 1 Level) and the SRA Achievement Tests (Level G) have only one each. Astronomy/Space and Meteorology are both represented on the achievement test batteries by five questions or less. Geology seems to have faired somewhat better. It must be emphasized that the number of test items reported in Table 2 represent entire academic disciplines--any one of which could be further broken down into broad topic areas.

Implications

If a school district annually administered the Comprehensive Tests of Basic Skills (Levels 3 & 4), the SRA Achievement Tests (Level F), or the Stanford Achievement Tests (Advanced Level) and had intended to use the scores earned on the Science subtest to evaluate the effectiveness of a new six-week unit on Oceanography being added to the Earth Science curriculum in the junior high school, they would find the students had apparently learned nothing because their scores would not have changed. The reason for this, of course, is that these standardized achievement test batteries contain no questions on Oceanography--hence, they are not sensitive to a change in the Earth Science curriculum related to Oceanography. Similarly, it would be extremely difficult to assess the effectiveness of a new or revised unit of Astronomy/Space or Meteorology since the Science subtest of the standardized achievement test batteries have,

at most, only five items related to these areas. Again, these tests are too short and, therefore, insensitive to the curriculum changes being assessed. The same argument can be made for the assessment of changes made in the Geology component in Earth Science.

Conclusion

While the purpose of this paper was to demonstrate the insensitivity of the Science portion of selected standardized achievement test batteries in assessing the effectiveness of curriculum changes in the Earth Sciences, the same may be advanced for their ability to evaluate curriculum changes in the Life and Physical Sciences. It is imperative, in these days of reduced resources within the local school district, to convince the pertinent decision makers that, despite their overall virtues, standardized achievement test batteries are not to be used to evaluate curriculum changes and that the time and money expended to impliment these changes justify the use of appropriate evaluation techniques and instrumentation. Only then can an accurate assessment of a curriculum's effectiveness be determined.

Table 1

Standardized Achievement Test Batteries:

Publisher, Publication Date, Level, Recommended Grades, and Item Breakdown

Test	Publisher	Pub. Date	Level	Recommended Grades	Number of Items			
					Earth Science	Life Science	Physical Science	Total
Comprehensive Tests of Basic Skills	CTB/McGraw Hill	1973	3	6.5-8.9	11	14	16	41
			4	8.5-12.9	3	21	16	40
Metropolitan Achievement Tests	Harcourt Brace Jovanovich, Inc.	1978	Adv. 1	7.0-9.9	16	22	17	55
SRA Achievement Tests	Science Research Associates, Inc.	1978	F	6.5-8.5	9	13	15	37(+3)*
		1978	G	8.5-10.5	10	9	16	35(+5)*
Sequential Tests of Educational Progress	Addison-Wesley Publishing Co.	1979	Adv. 1	7.5-10.5	25	25	25	75
Stanford Achievement Tests	Harcourt Brace Jovanovich, Inc.	1973	Adv.	7.0-9.9	9	25	26	60

*Some items within the Science portion of the SRA Achievement Tests did not fit into any of the three categories provided.

Table 2

Classification of Earth Science Test Items

into

Astronomy/Space, Geology, Meteorology, and Oceanography

Test	Level	Recommended Grades	Number of Items				
			Astronomy/ Space	Geology	Meteor- ology	Ocean- ography	Total
Comprehensive Tests of Basic Skills	3	6.5-8.9	0	7	4	0	11
	4	8.5-12.9	0	2	1	0	3
Metropolitan Achievement Tests	Adv. 1	7.0-9.9	3	8	4	1	16
SRA Achievement Tests	F	6.5-8.5	2	4	3	0	9
	G	8.5-10.5	4	3	2	1	10
Sequential Tests of Educational Progress	Adv. 1	7.5-10.5	5	12	5	3	25
Stanford Achievement Tests	Adv.	7.0-9.0	1	3	5	0	9